Evolutionary pattern in the *antR-cor* gene in the dwarf dogwood complex (*Cornus*, Cornaceae)

By: Chuanzhu Fan, Qiu-Yun (Jenny) Xiang, David L. Remington, Michael D. Purugganan, and Brian M. Wiegmann

Fan, C., Q.-Y. Xiang, D.L. Remington, M.D. Purugganan, and B.M. Wiegmann. 2007. Evolutionary pattern in the *antR-cor* gene in the dwarf dogwood complex (*Cornus*, Cornaceae). Genetica 130:19-37.

This is a post-peer-review, pre-copyedit version of an article published in *Genetica*. The final authenticated version is available online at: http://dx.doi.org/10.1007/s10709-006-0016-3

Abstract:

The evolutionary pattern of the *myc*-like anthocyanin regulatory gene *antR-Cor* was examined in the dwarf dogwood species complex (*Cornus* Subgenus *Arctocrania*) that contains two diploid species (*C. canadensis* and *C. suecica*), their putative hybrids with intermediate phenotypes, and a tetraploid derivative (*C. unalaschkensis*). Full-length sequences of this gene (~4 kb) were sequenced and characterized for 47 dwarf dogwood samples representing all taxa categories from 43 sites in the Pacific Northwest. Analysis of nucleotide diversity indicated departures from neutral evolution, due most likely to local population structure. Neighbor-joining and haplotype network analyses show that sequences from the tetraploid and diploid intermediates are much more strongly diverged from *C. suecica* than from *C. canadensis*, and that the intermediate phenotypes may represent an ancestral group to *C. canadensis* rather than interspecific hybrids. Seven amino acid mutations that are potentially linked to myc-like anthocyanin regulatory gene function correlate with petal colors differences that characterize the divergence between two diploid species and the tetraploid species in this complex. The evidence provides a working hypothesis for testing the role of the gene in speciation and its link to the petal coloration. Sequencing and analysis of additional nuclear genes will be necessary to resolve questions about the evolution of the dwarf dogwood complex.

Keywords: *Cornus* | Gene evolution | Hybridization | *Myc*-like anthocyanin regulatory gene | Nucleotide polymorphism | Polyploid | Speciation

Article:

Introduction

Identifying genetic changes at the DNA level underlying adaptive morphological divergence is essential to unraveling the molecular basis of speciation. Studies of the pattern of gene evolution, especially genes having a function associated with a key morphological trait potentially involved in species divergence, would be particularly illuminating in this regard. Such studies in a hybrid–polyploid complex may further elucidate how gene evolution in the hybrids and polyploidy translate into novel phenotype due to genome dynamics associated with gene and genome duplication (see reviews by Wendel 2000; Wolfe 2001; Moore and Purugganan 2005). Some recent studies have proposed that regulatory gene evolution can be a significant factor in organismal diversification (Wilson 1975; King and Wilson 1975; Dickinson 1988; Doebley 1993). Changes in regulatory loci, for example, are thought to underlie the evolution of some developmental mechanisms that result in morphological differentiation between taxa (Carroll 1995; Palopoli and Patel 1996; Purugganan 1998, 2000). Studies also suggest that regulatory genes in hybrids also play a key role in the processes of generating novel morphological and physiological phenotypes acted on by natural selection (e.g. Doebley and Lukens 1998; Purugganan 1998; Wendel 2000; Kellogg 2002; Simpson 2002; Levine and Tjian 2003; Papp et al. 2003).

The dwarf dogwood or bunchberry (*Cornus* Subg. *Arctocrania* Endl. Ex Reichenb.) species complex provides a nice system to study gene evolution associated with species divergence and hybrid novelty. These plants are perennial rhizomatous ground cover herbs. Three species, *C. canadensis* L., *C. suecica* L., and *C. unalaschkensis* Ledeb., were described from the group. The first two species are diploid (2n = 22), and the third is a tetraploid (2n = 44) (Bain and Denford 1979). The three species are mainly distributed in areas of high latitudes and elevations of the Northern Hemisphere. *Cornus canadensis* occurs from northeastern North America westward to northeastern Asia, extending southward to the Rocky Mountains, Appalachian Mountains, and mountains of Japan. *Cornus suecica* has more northerly and coastal distribution in North America and Eurasia (Murrell 1994). The tetraploid, *C. unalaschkensis*, is restricted to the Pacific region of North America where the ranges of the two diploid species overlap. The two diploid species are easily distinguished by their differences in petal color and leaf morphology. *Cornus suecica* has dark purple petals and several pairs of chlorophyllous sessile leaves with parallel veins, whereas *C. canadensis* has creamy white petals and a cluster of green leaves only on the uppermost node, and the leaves are shortly petiolate with pinnate venation (Table 1). In the Pacific Northwest and northeastern North America, the two species overlap and hybridize, resulting in a species complex consisting of two morphological extremes (corresponding to the two diploid species) and a wide array of intermediate forms combining different morphological features of the two morphological extremes (Bain and Denford 1979; Murrell 1994). These intermediate forms have white, purple, or bicolored petals variable in the relative portion of the purple color, from the tip to nearly entirely purple except the base center of the petal. The tetraploid species *C. unalaschkensis*, endemic to the Pacific Northwest, is among the bicolor intermediate forms, characterized by the apical half purple and basal half cream (Murrell 1994; see Table 1). Based on its morphological intermediacy, the tetraploid species was considered to derive from hybridization between *C. canadensis* and *C. suecica* (Bain and Denford 1979). In the Pacific Northwest, the bicolored floral phenotype is the most common and occurs across the distribution of both parental species. For example, *Cornus suecica* occur in completely open areas of high latitude and high elevations (from the transitional zone to above tree line) where wind is strong and UV light is intense. These plants are short and develop dark purple petals as

well as normal, paired green leaves at the upper three nodes of the stem. These characters are considered to be a potential adaptation to the strong light and UV environment (Stapleton 1992). In contrast, *C. canadensis* occur in coniferous forests of lower elevations and produce white petals, a ‘‘whorl’’ of larger green leaves only on the uppermost node of the stem as an adaptation to an environment with lower light intensity. The bicolor-flowered tetraploid *C. unalaschkensis* and diploid intermediates produce a whorl of large green leaves on the upper most nodes as well as a pair of reduced green or scaly leaves at the second and even third nodes from the stem apex (Table 1). These plants were found in both habitats across the elevational ranges inhabited by the two parental species.

Table 1. Morphological identification for four groups of dwarf dogwoods

Taxa	Leaves at node 2	Leaves at node 3	Leaves at node 4	Branches at node 1	Petal color	Hypanthium	Leave petiole	Leave midvein
C. canadensis (CC)	Scale	Scale	Scale	Reduced	Cream	Covered with hairs	Present	Present
C. canadensis > *C. suecica* (Group CH)			Scale	Reduced	Cream or purple-tipped	Covered with hairs	Present	Present
C. unalaschkensis (CU)			Scale	Reduced	Bicolor-apical half purple, basal half cream	Covered with hairs	Present	Present
C. suecica > C. *canadensis* (Group CH)	Green	Green	Green	Not reduced	Purple with a cream base	Lower half covered with hair	Present	Present or absent
C. suecica (CS)	Green	Green	Green	Not reduced	Purple	Hairs only at base	Absent	Absent

It is known that anthocyanin pigments are largely responsible for pink, red, purple and blue coloration in plant tissues, especially in flowers (Mol et al. 1998). Abundant evidence indicates that an important function of the pigments is protecting flowers from damage by the UV light (Stapleton 1992). The acquisition of purple petals in *C. suecica* (none of the other dogwood species has purple petals or occurs in the same habitat) was probably important to the species for adaptation to habitat at higher elevation with more intense UV light as well as divergence from the white flowered *C. canadensis*. The coexistence of both white and purple colors in the petals of the hybrid and polyploid individuals is a novelty and potentially a key trait permitting their occurrence under both strong and lower UV light conditions. Given the pattern of petal variation associated with ecological difference and species divergence, it is interest to investigate the evolutionary pattern of genes regulating the anthocyanin pathways in the group. Various studies have indicated that the expression of anthocyanin regulatory genes in *Zea mays, Nicotiana, Arabidopsis, Petunia, Antirrhinum majus, Gossypium, Perilla*, and *Solanum lycopersicum* activate the structural genes and induce pigmentation in a wide variety of tissues, especially in flowers, and that mutations in a case of *myc*-like regulatory genes result in partial expression of anthocyanin pigments in petals (Ludwig and Wessler 1990; Radicella et al. 1991; Martin et al. 1991; Goodrich et al. 1992; Consonni et al. 1992, 1993; Lioyd et al. 1992; Quattrocchio et al. 1993, 1998; Goldsbourough et al. 1996; Hu et al. 1996; De-Vetten et al. 1997; Mol et al. 1998; Gong et al. 1999). Evidence from *Arabidopsis, Zea mays, Brassica*, and *Drosophila* demonstrated that regulatory genes can harbor significant variation responsible for adaptive morphological evolution and sometimes can also exhibit low molecular diversity due to strong selection (review in Purugganan 2000). For example, in *Arabidopsis*, three floral developmental genes (CAL, AP3, and PI) show accelerated protein polymorphism linked to positive selective pressure (Riechmann and Meyerowitz 1997; Purugganan and Suddith 1998, 1999). In the catostomid family fishes, protein polymorphisms associated with reproductive morphological

diversification following hybridization and polyploidization were demonstrated (Ferris and Whitt 1979). In *Brassica*, the *BoCal* gene has low sequence variation, but positive selection was detected (Purugannan 2000).

We have previously reported on the isolation and molecular evolutionary analysis of a *myc*-like anthocyanin regulatory gene in the dogwood genus *Cornus* (*antR-Cor* hereafter) and described its utility for resolving phylogenetic relationships within the genus (Fan et al. 2004). In the present study, we examine the pattern of genomic sequence variation of *antR-Cor* (~4,000 bp) for the species complex from the Pacific Northwest of North America and further investigate the processes that may have shaped the sequence variation in *antR-Cor*. Finally, we use the *antR-Cor* sequence data to test assumptions about the evolution of the dwarf dogwood complex, including the hybrid origin of the bicolored phenotypes and the allopolyploid origin of the tetraploid species. The evidence provides a working hypothesis for testing the role of the gene in speciation and its link to the petal coloration. Our results suggest an association of amino acid substitutions at some sites with petal color phenotypes, and raise new questions about the origins of the bicolored phenotypes and *C. unalaschkensis*.

Table 2. The dwarf dogwood samples and outgroups used in this study

Groups	**Total number of populations**	**Voucher of collection: population #**	**Color of petals**	**Collection localities**
CC (*C. canadensis*)	7	6, 7, 8, 17, 18, 21, 40	White	AK, BC, YK
CH (hybrids)	12	14, 15, 16, 19, 20, 29-1, 29-2, 32, 33, 30, 41, 42	Bicolor (32, 33, 30, 41, 42), purple (29) or White	AK, BC, YK
CU (*C. unalaschkensis*)	23	1, 2, 3, 4, 5, 9, 10, 11, 12, 13, 22, 23, 24, 25, 26, 28, 31, 34, 35, 36, 37, 38, 39	Bicolor	ID, WA, OR, BC, AK, YK
CS (*C. suecica*)	5	27-1, 27-2, 43-1, 43-2, 94-388	Purple	AK, Norway
Outgroups	1	*C. florida* (02-16)	—	Mexico

Note: For voucher of collection, the first number represent the population and the second number indicates the individual

Materials and methods

Sampling

Samples of dwarf dogwoods were collected from 43 sites across the Pacific Northwest region of North America where all morphological forms and ecotypes co-occur. Plant samples were classified into three species (*C. canadensis, C. suecica*, and *C. unalaschkensis*) and hybrids based on a combination of morphological diagnostic characters provided in Bain and Denford (1979) and Murrell (1994) from cytological and morphological analyses (Table 1): Group ‘‘CC’’—(*C. canadensis*); Group ‘‘CS’’—(*C. suecica*); Group ‘‘CU’’—(*C. unalaschkensis*); and Group ‘‘CH’’—(diploid intermediates representing putative hybrids). Group ‘‘CC’’ includes six samples collected from various sites of the collecting area; group ‘‘CS’’ contains seven samples from Alaska and one additional sample (#94-388) from Norway; *C. unalaschkensis* (‘‘CU’’) includes 23 samples from across a wide range of the collecting areas. The intermediates, group ‘‘CH’’, have 10 samples from various collecting sites (Table 2). The sampling sites span geographical areas of Idaho (ID, USA), Oregon (OR, USA), Washington (WA, USA), Alaska (AK, USA), British Columbia (BC, Canada), and the Yukon (YK, Canada) (see Fig. 1). One samples representing a species of *Cornus* from the sister clade of dwarf dogwoods, *C. florida*

(Fan and Xiang 2001; Xiang et al. 2006) was included in the analyses to provide outgroups for rooting the phylogenetic trees.

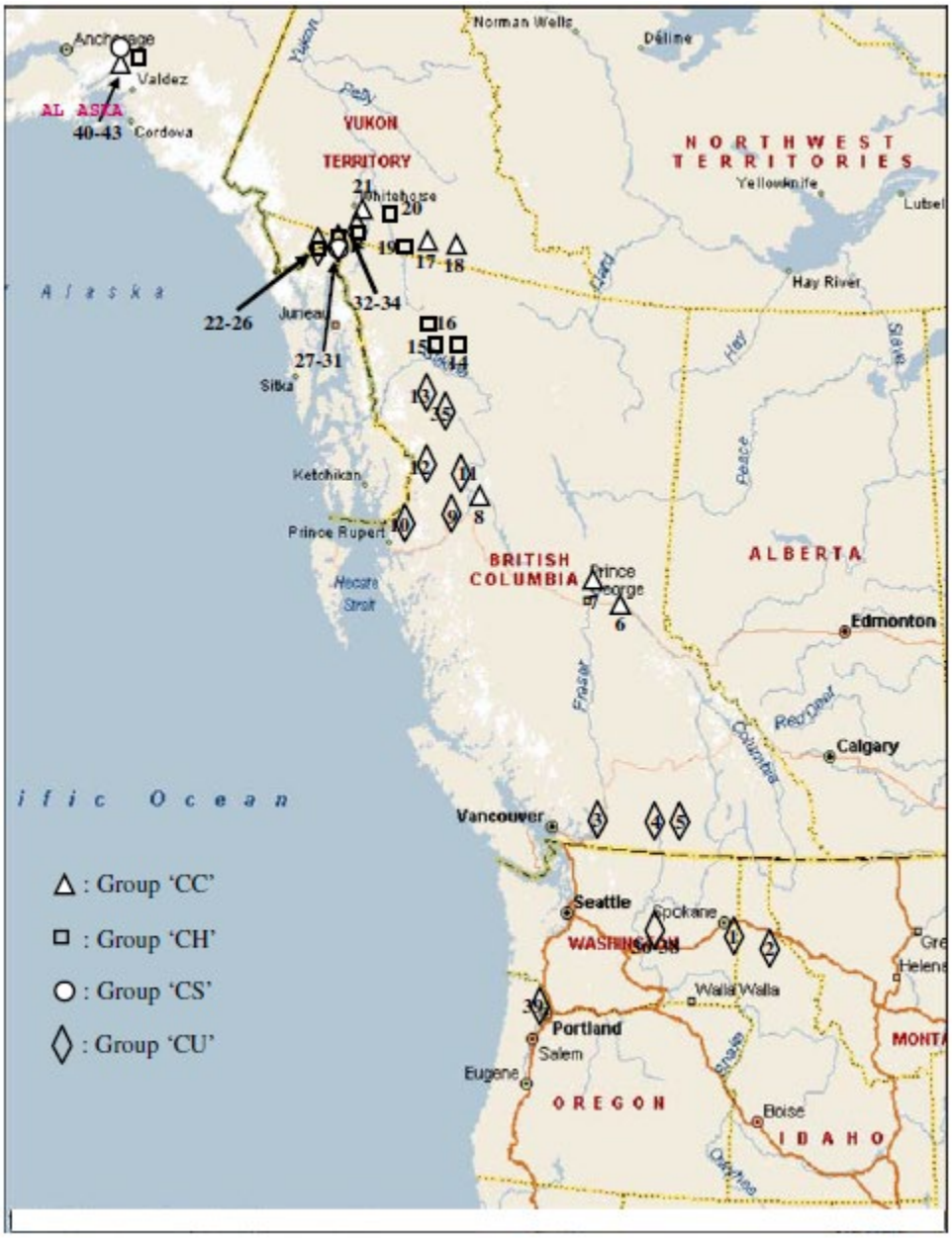

Figure 1. Dwarf dogwood sampling locality. Symbols (square, triangle, diamond or circle) represent the approximate location of sampled population. The number adjacent to each symbol is a unique population identifier

DNA extraction, PCR amplification, and sequencing

Total genomic DNA was extracted from fresh leaves, using a previously described protocol (see Xiang et al. 1998). The genomic sequence of the *myc*-like anthocyanin regulatory gene was amplified using the following primer combinations: F0A-R2A2, F2A (or F2A1)-R3′, F4A-R4A, F6A-R7A, and F7A2-R9A (Fig. 2; Fan et al. 2004). PCR reactions contained the following combinations: 5 µl of 10 × Mg^{2}+ free buffer, 6 µl of 25 mmol/l $MgCl_2$, 6–10 µl of 2.5 mmol/l

dNTPs, 0.5 µl of 20 µmol/l forward primer, 0.5 µl of 20 µmol/l reverse primer, 5 µl of DMSO (dimethyl sulfoxide), 1–5 µl of Bovine serum albumin (BSA, 10 mg/ml), 0.3 µl of *Taq* polymerase (Promega), 5–10 µl of 20 ng/µl total DNA extract, and calibrated to final 50 µl using deionized water. Amplification conditions were as follows: (1) 94°C for 30 s for one cycle; (2) 30–40 cycles of 94°C for 45 s, 50–60°C for 1 min, 72°C for 1.5–2.5 min; (3) a terminal phase at 72°C for 6 min. The process of sequencing purified PCR products is described in Fan et al. (2004). Additionally, TA cloning of PCR products was used to obtain sequences from a few CU samples to confirm the heterozygosity for insertion–deletion (indel) polymorphisms. The sequences from cloning products were identified as only two alleles, one with indel sequences and one without indel sequences. This also confirms that the antR-Cor is a single copy in both tetraploids and diploids. Therefore, heterozygosity of the indel in the tetraploids may simply be due to the presence of four alleles, reducing the likelihood of homozygosity. The sequence chromatogram output files for all samples were checked visually and edited before alignment. The sequences were aligned using Clustal X (Thompson et al. 1997), and adjusted manually. Presence of another copy of antR-Cor in Cornus is possible. Initial PCR and sequencing using the degenerate primers designed from GenBank sequences of other taxa detected two divergent types of sequences. The sequences included in this analysis are all from the same copy amplified by copy-specific PCR primers described above. Sequence comparison indicates that this copy is homologous to the myc-like anthocyanin regulatory gene identified in other model organisms (see Fan et al. 2004).

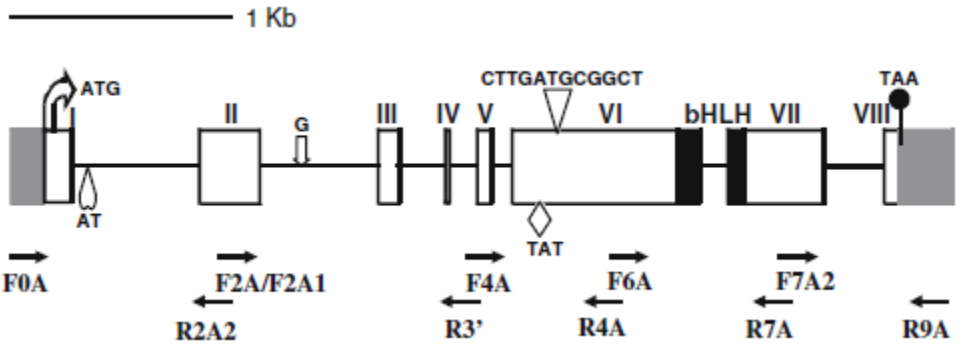

Figure 2. Schematic map showing the overall structure of *antR-Cor* gene for dwarf dogwoods as deduced from the genomic sequences. The position of the ATG translation start and the TAA translation stop codons are indicated. The region encoding bHLH is shown in black. The flanking regions are shown in shadow. Four indels are indicated by heart (2 bp), arrow (1 bp), diamond (3 bp), and triangle (12 bp), respectively. The boxes represent exons, and the line represents introns. Exons are ordered as I–VIII. The relative position of primers used for PCR amplification is also shown. The sequences of these primers are as followed: F0A (TCACTGAGTGGGTGTCTTAAG); R2A2 (CCACTCCGTATCCGTGAGGT); F2A (TTTATGAGTCCCTTGYGGTCAC); F2A1 (GTTCAGGCGGTCGAATTCAATGCCG); R3′ (CCSAGCTCAAYYACWCCTCC); F4A (GCGATATTGCCATTTGTCTG); R4A (CATTTATGGAAGTAAGGTCCC); F6A (CTGACCTCGTTGGACCTTC); R7A (CAAGCAACAAGCGCTCCCT); F7A2 (GAGCTGGAGATCAACCTCG); R9A (CTATCCACAAGAAACACYTGC)

Test of sequence diversity and neutrality

To investigate sequence diversity within the dwarf dogwood species complex, patterns and rates of sequence variation were assessed within each group (‘‘CC’’, ‘‘CS’’, ‘‘CU’’, and ‘‘CH’’). Sequence divergence between groups and tests of deviation from neutral equilibrium expectations were conducted using methods in the program DnaSP (Rozas et al. 2003). Sequence

diversity tests were performed using nucleotide diversity (π) (Nei 1987) and the population mutation parameter of Watterson's θ (denote here as θ_w) (Watterson 1975) in DnaSP version 4.0 (Rozas et al. 2003). Tests of deviation from neutrality were conducted using Tajima's (1989) and Fu and Li's (1993) methods in DnaSP. For Fu & Li's test, we performed the analyses in two ways, with and without outgroups for a comparison. Tajima's and Fu and Li's tests of selection examine deviation from neutral expectations by estimating the test statistic D for θ. Both methods are statistical analyses of haplotypes (chromosome). Two alleles (or haplotypes) for each sample were used in the tests for all the groups (CC, CS, CH, and CU). Method for inferring haplotype sequences is detailed in ''Results'' below.

Phylogenetic and gene genealogical analyses

To understand the genealogical history of the gene and the study group, we performed phylogenetic analysis and constructed a haplotype network of allele sequences. Phylogenetic analysis of the DNA sequences from haplotypes in all samples was performed using the neighbor-joining method with Jukes–Cantor distance implemented in PAUP* 4.0b10 (Swofford 2002). Phylogenetic trees were rooted using *C. florida*, the sister group of the dwarf dogwoods, and clade support was estimated by bootstrap analysis of 1,000 replicates (Felsenstein 1985). The haplotype network was constructed using the method of Templeton, Crandall, and Sing (TCS, Templeton et al. 1992; Clement et al. 2000). The program uses a 95% statistical parsimony support criterion and takes into account population-level phenomena (e.g., hybridization, gene flow, and recombination) in estimating a network of relationships (Templeton et al. 1992; Clement et al. 2000). The default setting of the program allows 24 steps as the maximum number of mutations for connecting two haplotypes. The analysis with this setting prevented the connections of a few divergent haplotypes to the network. We therefore increased the number of allowable steps in the TCS analysis to 100 steps.

Results

Sequence data

The *antR-Cor* sequences generated for the 47 samples of dwarf dogwoods varied from 4,023 to 4,040 bp in length, except for two with incomplete sequences (samples #35 and 40). The sequences for all samples have been deposited in GenBank as accession number of AF465415–AF465418 and AF 493694–AF493736. This gene in dwarf dogwoods contains eight exons and seven introns. The length of the coding region is 1863, 1875, or 1878 base pairs, and the seven introns have a total of 1864 or 1867 bp. In addition to nucleotide substitutions, four indels were found, of which two are from exon VI, one from intron 1, and one from intron 2 (Fig. 2). For the exon indels, three base pairs (TAT) are mainly found in intermediates and tetraploids and the 12-bp indel in exon VI (Fig. 2) is present in ''CS'' and nearly all ''CH''. All ''CU'' individuals contained alleles both with and without the 12-bp insertion but otherwise show high similarity. All other regions of the sequence for the two alleles in the hybrids and tetraploids are identical except at several polymorphic sites at which chromatograms showed the presence of two nucleotides. The presence of indel variants in a single individual was detected by observing ''clean'' (non-polymorphic) sequences until the indel position where sequences from both alleles were mixed and a single base pattern could not be called. The heterozygotic condition at the

indel was confirmed by cloning and sequencing of both alleles in several samples showing unreadable sequences starting at the indel sites. Haplotypes were inferred at the polymorphic sites by observing which nucleotide was most common in individuals homozygous for the 12 bp insertion and deletion, respectively, and assigning that nucleotide to the insertion and deletion alleles, respectively in the individuals carrying both versions. This process for inferring haplotypes introduces some bias into analyses of gene genealogy. However, unambiguous demarcation of haplotypes would not be possible even with complete sequencing of clones of PCR products, because multiple PCR products were required to amplify the whole gene and because cloning could create chimeric haplotypes due to the possible occurrence of PCR-generated recombinants. Therefore, we split the heterozygous indels and polymorphic sites into two alleles for later sequence diversity and phylogenetic analyses. The full data matrix contains 738 variable sites (19.2%) and 171 (4.46%) parsimony-informative sites within the 90 dwarf dogwood allele sequences.

Table 3. Molecular diversity in the myc-like anthocyanin regulatory gene in the dwarf dogwoods

Group	Length	*n*	N_{hap}	*S*	π_{all}	π_{silent}	$\pi_{non\text{-}code}$	π_{synon}	$\pi_{non\text{-}syn}$	θ_w ($4N\mu$)	Tajima's *D*	Fu and Li's *D**
CC	3,793	12	9	29	0.00228	0.00229	0.00210	0.00316	0.00226	0.00245	-0.31461	0.80419
											$P > 0.1$	$P > 0.1$
CS	3,998	10	6	64	0.00652	0.00501	0.00432	0.00856	0.00915	0.00548	0.93045	1.63903
											$P > 0.1$	$P < 0.02$*
CH	3,805	24	22	90	0.00625	0.00700	0.00733	0.00536	0.00501	0.00635	-0.05841	1.30509
											$P > 0.1$	$P < 0.05$*
CU	3,891	44	40	70	0.00372	0.00386	0.00405	0.00287	0.00319	0.00347	-0.39092	1.75670
											$P > 0.1$	$P < 0.02$*

CC—*C. canadensis*; CS—*C. suecica*; CH—Hybrids; CU—*C. unalaschkensis*. *n*: sample size; N_{hap}: the number of observed haplotypes; S: the number of observed mutation sites. π_{all}: nucleotide diversity at all sites; π_{silent}: nucleotide diversity at silent sites (synonymous and non-coding regions); $\pi_{non\text{-}code}$: nucleotide diversity at non-coding regions (franking region and introns); π_{synon}: nucleotide diversity at synonymous sites; $\pi_{non\text{-}syn}$: nucleotide diversity at non-synonymous sites; θ_w: Watterson's estimate of mutation parameter. *Significant at the $P < 0.05$ level; **significant at the $P < 0.01$ level

Nucleotide diversity and neutrality test

The level of sequence variation observed in the four ''species'' groups is substantially different (Table 3). The total polymorphic sites are shown in Fig. 3. Sequence analyses indicated that Group ''CC'' contains 29 segregating sites and each has a distinct haplotype. This group has the lowest sequence diversity among all π values (including all sites and different types of sites) and lowest θ_w value (Table 3). Among 29 mutation sites across the entire sequence within this group, 14 are from protein coding regions. For polymorphic sites in exons, 10 are non-synonymous mutations and four are silent mutations. Group ''CS'' has much higher sequence variation and diversity than Group ''CC'' as measured by π and θ_w values, which are approximately three times the values of those for group ''CC'' in all tests (Table 3). ''CS'' possess six haplotypes with 64 segregating sites. Forty-one of 64 segregating sites are from coding regions. Among 41 polymorphic sites in coding regions, 32 are replacement mutations. Group ''CH'' has nearly the level of diversity that CS has, while CU is intermediate (Table 3). A total of 90 mutation sites were found in Group ''CH'', of which 35 are from coding regions. Twenty-six of the 35 sites in coding regions are replacement mutations, and eight are silent mutations. In group ''CU'', 70 mutation sites were detected. Among them, 34 mutations are from coding regions, and 28 of 34 mutations are non-synonymous. The sequence divergence test shows that groups ''CH'' and

‘‘CU’’ are more similar to ‘‘CC’’ than to ‘‘CS’’, and ‘‘CH’’ and ‘‘CU’’ are similar to each other (Table 4).

Table 4. Sequence divergence between groups

Interspecific	K_{all}	K_{silent}	$K_{non\text{-}code}$	K_s	K_a	K_a/K_s
CC–CH	0.00516	0.00536	0.00544	0.00499	0.00484	0.971
CS–CH	0.01254	0.01324	0.01405	0.00946	0.01144	1.212
CC–CS	0.01515	0.01575	0.01667	0.01149	0.01421	1.239
CC–CU	0.00539	0.00524	0.00522	0.00534	0.00570	1.068
CS–CU	0.01263	0.01316	0.01336	0.01223	0.01179	0.964
CH–CU	0.00549	0.00596	0.00630	0.00438	0.00474	1.082

K_{all}: the average number of nucleotide substitutions per site between groups for all sites; K_{silent}: the average number of nucleotide substitutions per site between groups for silent sites (synonymous and non-coding regions); $K_{non\text{-}code}$: the average number of nucleotide substitutions per site between groups for non-coding sites (flanking region and introns); K_s: the average number of nucleotide substitutions per site between groups for synonymous sites; K_a: the average number of nucleotide substitutions per site between groups for non-synonymous sites

All neutrality hypothesis tests examined in this study yielded non-significant deviation from neutrality in Tajima’s *D*, but both Fu and Li’s tests yielded positive *D* value that are significantly different from neutrality in CS, CH, and CU groups (Table 3). Results of these tests in the CU group must be interpreted with caution because the sampled alleles may represent homeologous loci in the tetraploid *C. unalaschkensis* rather than alleles, depending on the nature of chromosome pairing in the tetraploids, but it is noteworthy that the *D* values do not differ greatly from the range of values observed in the other groups. Futhermore, the locus behaves as a single-copy locus in the tetraploid, with only minor differences between copies, which suggests random pairing of the four sets of homologous chromosomes rather than the presence of distinct homologs. The ratios of non-synonymous to synonymous nucleotide diversity both within and between groups is around one (Table 4), indicating a near-absence of functional constraint in the *antR-Cor* gene in the dwarf dogwood complex. These results contrast with those obtained for the genus *Cornus* as a whole, in which the *antR-Cor* gene showed a moderate level of selective constraint ($K_a/K_s = 0.41$; Fan et al. 2004).

Gene genealogy and population subdivision

Clustering of sequences using neighbor-joining identified a strongly supported cluster consisting of the purple/high-elevation ‘‘CS’’ samples as well as two ‘‘CH’’ samples with purple petals (#29), and a weakly supported cluster consisting of the ‘‘CC’’, ‘‘CU’’, and remaining ‘‘CH’’ samples. Within the latter cluster, a smaller group closely corresponds to the white/low-elevation Group ‘‘CC’’ and samples from the ‘‘CH’’ group with white petals (Fig. 4). This grouping is congruent with the genetic distance estimation showing a smaller distance between the ‘‘CC’’ and ‘‘CH’’, ‘‘CU’’ than between the ‘‘CS’’ to any of them.

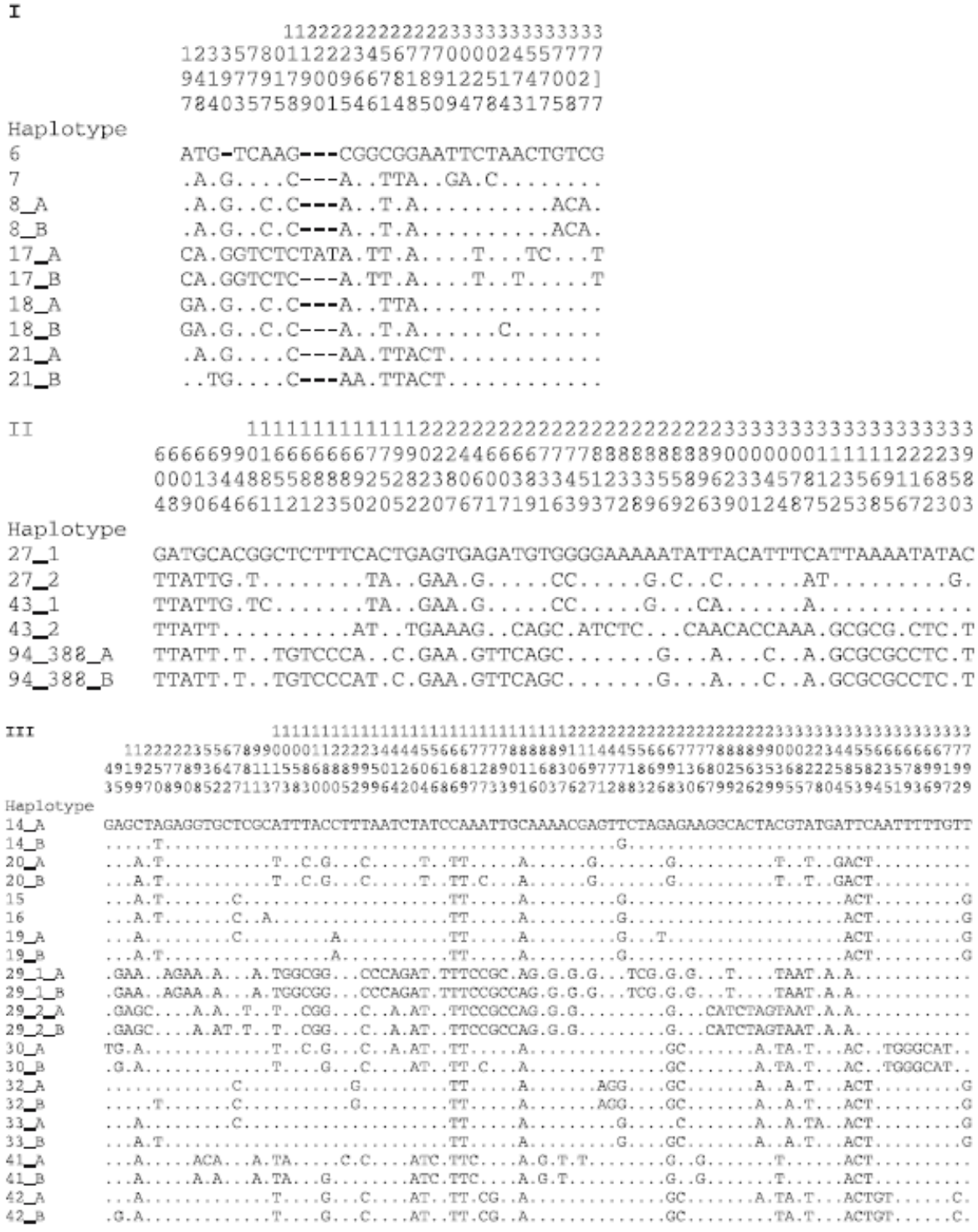

Figure 3. Polymorphic sites of *antR-Cor* of haplotypes in four groups. ‘‘A’’ and ‘‘B’’ are designated as two alleles if a samples has two haplotypes. Dots indicate identity to topmost sequence. Dashes represent gaps. Numbers following the population identification number represent individuals. I: Group ‘‘CC’’; II: Group ‘‘CS’’; III: Group ‘‘CH’’; IV: Group ‘‘CU’’

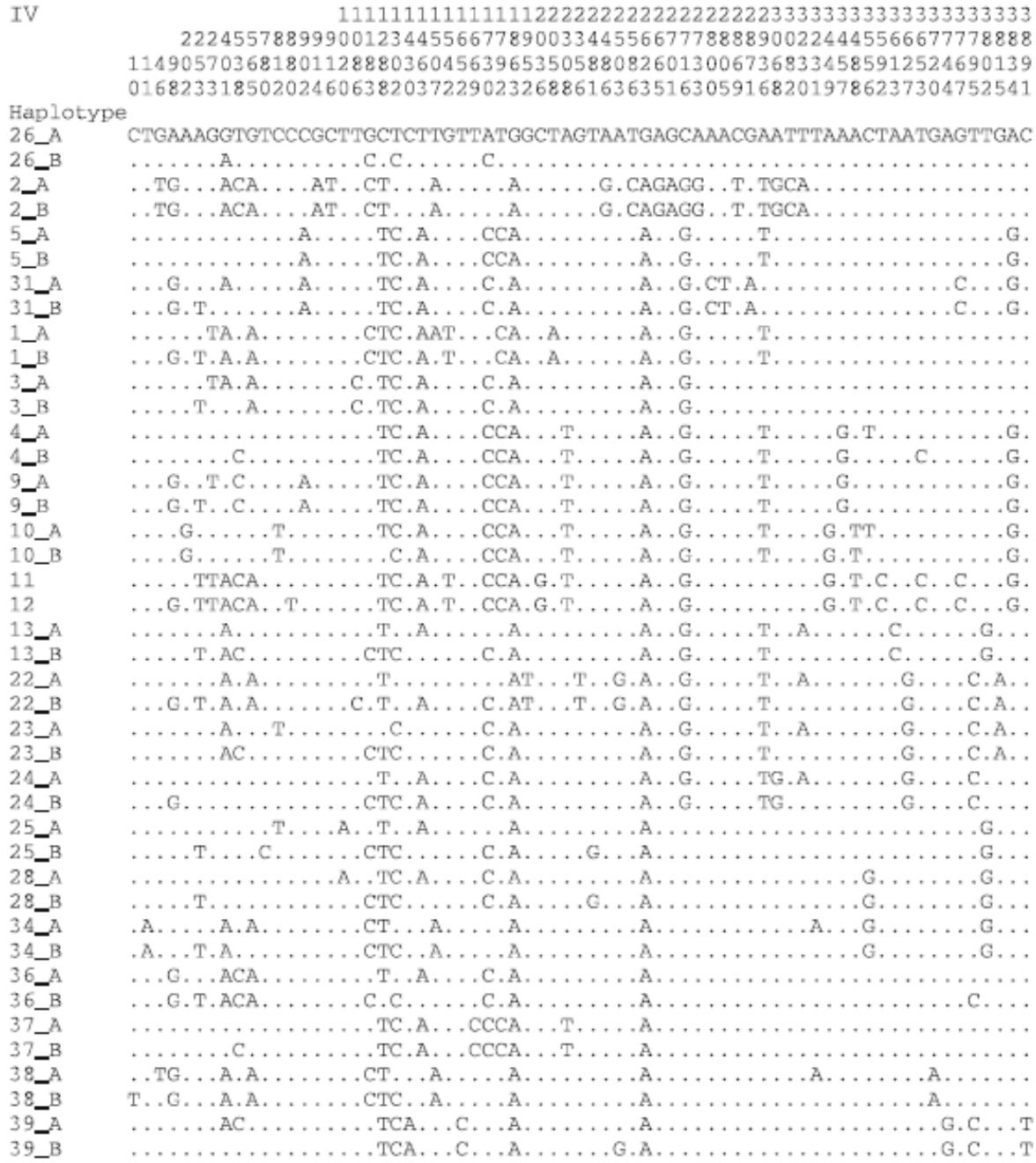

Figure 3, continued

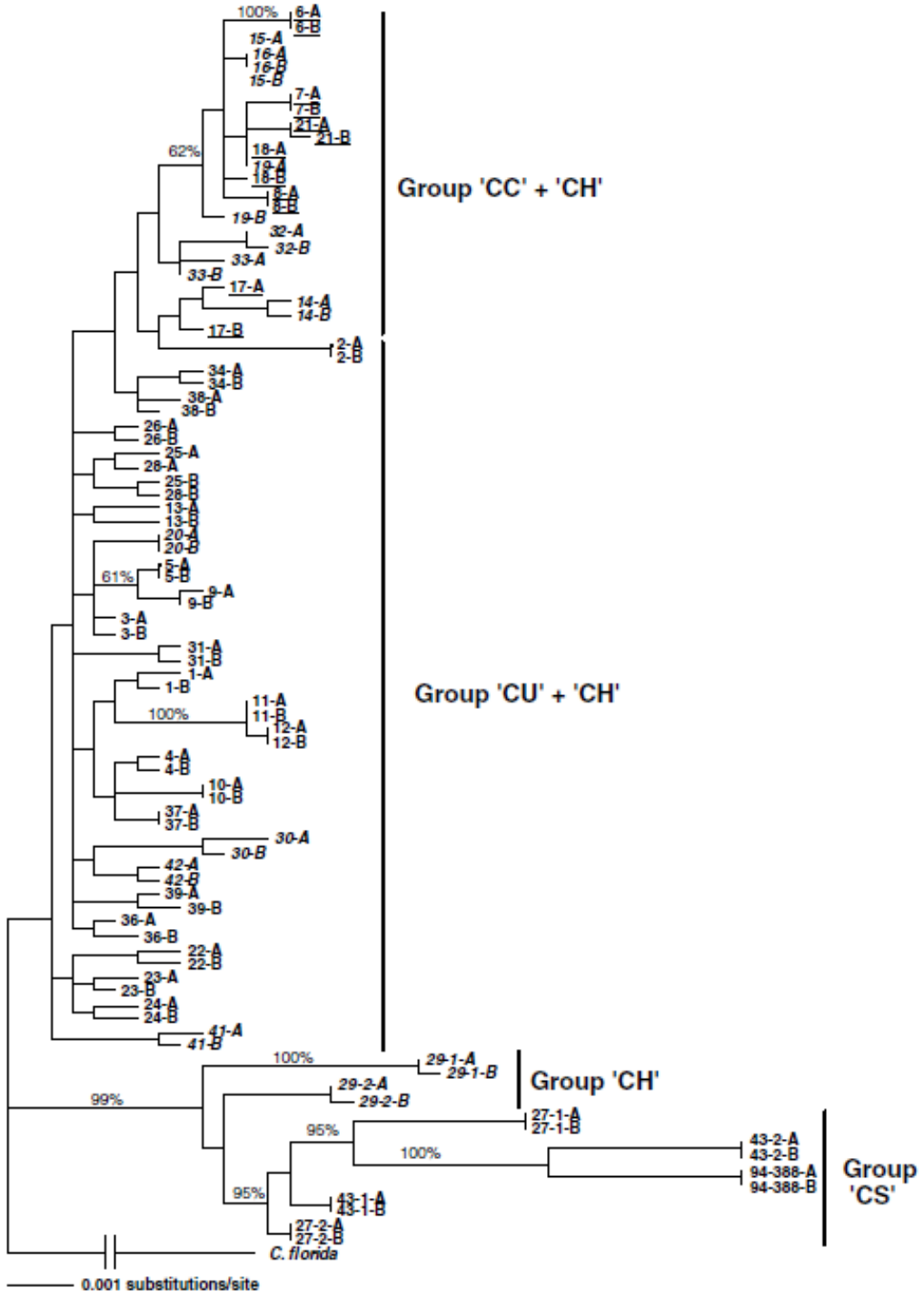

Figure 4. Neighbor-joining tree of *antR-Cor* gene sequences. Each sample contains two alleles (donated as ‘‘A’’ and ‘‘B’’). Bootstrap support values > 50% are shown above branches. Samples from CC group are underlined, and samples from CH group are italicized

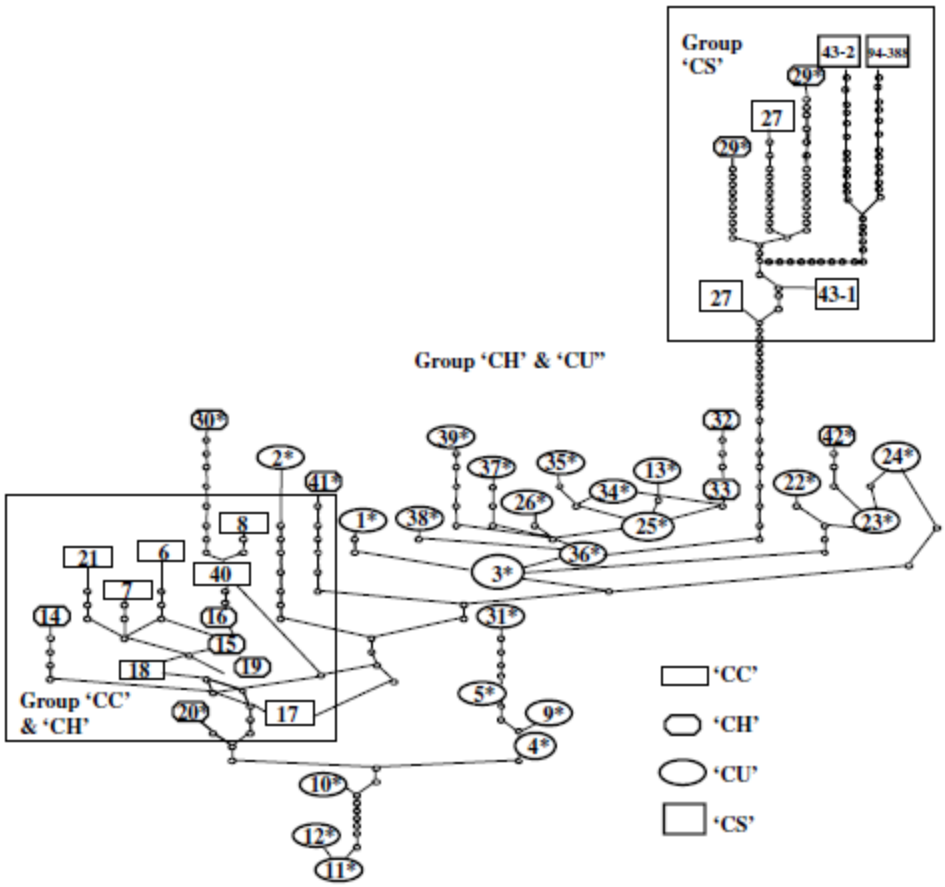

Figure 5. Statistical parsimony haplotype network of *antR-Cor* gene in the dwarf dogwoods. Small circles represent missing haplotypes with > 95% statistical parsimony support. An asterisk ‘‘*’’ indicates the heterozygous state of the 12-bp indel (see Fig. 2). Because the two alleles from each individual are grouped together in the neighbor-joining analysis (Fig. 4), only a single consensus sequence from each individual was used in the network reconstruction for simplicity

Figure 5 shows the 95% statistical parsimony-based haplotype network for the 46 samples. Since alleles from the same sample (or individual) are always grouped together in phylogenetic analysis, thus only a single consensus sequence of the two alleles from each sample (i.e., a single consensus sequence of the two alleles from each sample was included in the network analysis, using standard degeneracy codes at heterozygous sites) was included in the network analysis for simplicity (Fig. 5). The first clade includes the five haplotypes from Group ‘‘CS’’ and two samples from a single ‘‘CH’’ sample (#29) bearing completely purple petals. This clade has many fixed sequence variants that are not found in other haplotypes (Fig. 5). The second clade consists of the remaining haplotypes, which are connected via a complex network (Fig. 5). Six regions with non-linear connections among haplotypes were found in the TCS network, suggesting gene recombination. The network also recovers a closer relationship of the ‘‘CH’’ and ‘‘CU’’ haplotypes to those of the ‘‘CC’’, consistent with the clustering analysis. All haplotypes in CC and CH individuals with white petals (in box labeled ‘‘group CC and CH’’ in Fig. 5) are derived from a single node in the network, but a few of CU and bicolored CH haplotypes are also part of this clade. Visual inspection of the sequence alignment in Fig. 6 reveals possible additional recombination events not identified by TCS. Haplotype 30 with the 12-nucleotide insertion is nested in the clade missing the insertion (Fig. 5), while haplotypes 32

and 33 lacking the 12 bp are nested among the samples all having the 12 bp. These haplotypes appear to have arisen by recombination, producing results incongruent with the neighbor-joining analysis (Fig. 4) in which all haplotypes without the 12 bp-insertion are grouped in the same cluster. Apparent recombination within the haplotype network could also result from incorrect inference of haplotypes from sequence polymorphisms. The root position could not be evaluated for the haplotype network, but the large number of steps separating the ''CS'' clade from the remaining sequences is consistent with a root position between these clades, as suggested by the neighbor-joining analysis. It further indicates that two ''CU'' haplotypes #36 (Wenatchee Mts, Washington, USA) and #3 (northern Idaho, USA) are ancestral among the ''CU'' haplotypes (based on their internal positions and connecting with many other haplotypes). These haplotypes give rise to many other closely related ''CU'' haplotypes with few mutational steps, suggesting recent expansion of the ''CU'' group.

Associations between amino acid sequence variation and petal color

We translated DNA sequences of each haplotype for 47 samples into amino acid sequences. Eight of 47 samples have two haplotypes with only one amino acid difference (Fig. 6). Visual examination of variable sites in amino acid sequences show that variations at seven sites are correlated with petal colors (Site 19, 228, 307, 380, 436, 464, 518) (Fig. 6). Site 19 from the interaction domain and site 228 from the acidic domain, are fixed with QV in the purple phenotype and with RA in the white and bicolor phenotype, respectively. Sites 380 from the acidic domain and site 518 from the C-terminal domain, are fixed with DM in most white phenotypes and GL in the purple and most bicolor phenotypes (Fig. 6). Site 464 from the bHLH domain is fixed with V in the white colored phenotypes and with I in the purple and bicolor phenotypes (Fig. 6). The amino acid changes involve substitutions between arginine (R) and glutamine (Q) at site 19, alanine (A) and valine (V) at site 228, aspartic acid (D) and glycine (G) at site 380, serine (S) and leucine (L) at site 436, valine (V) and isoleucine (I) at site 464, and methionine (M) and leucine (L) at site 518 (Fig. 6). Amino acid haplotypes at these seven sites show that RAADSVM (10 of 11 samples) and RASDSVM (1 of 11 samples) are associated with white petals; RAAGSIL (2/29), RASGSIL (23/29), RASDLVM (1/29), RASGSVL (2/29), and RASGSVM (1/29) are associated with bicolored petals including the two samples with red petals; QVSGLIL (5/7) and QVSGSIL (2/7) are associated with purple petals (Fig. 6).

Discussion

Gene evolution, genealogy and dwarf dogwood species limits

The pattern of sequence variation of the *antR-Cor* gene in dwarf dogwoods shows significant positive deviation from neutrality under Fu and Li's test in all groups other than CC. By contrast, Tajima's test is slightly negative in all groups except CS, and does not deviate significantly from zero. This discrepancy appears to be due to the high similarity between the two alleles carried by all individuals. As a consequence, virtually all polymorphisms are present in at least two alleles, resulting in an almost complete absence of singletons whose frequency forms the basis for the Fu & Li's test (Fu and Li 1993). The non-neutral evolution of the gene was also suggested by our previous analyses of this gene for 10 divergent dogwood species (Fan et al. 2004). The lack of allelic variation within individuals could be due either to selfing or to biparental inbreeding in

small isolated local populations. Although other subgroups of *Cornus* have been found to be self-incompatible (Gunatilleke and Gunatilleke 1984), preliminary inflorescence bagging experiments in greenhouse suggest self-compatible in the dwarf dogwoods (Xiang unpublished). The rhizomatous nature of the dwarf dogwoods could also result in local populations composed largely of clones, resulting in small effective population sizes and elimination of heterozygosity at most loci in a relatively small number of generations. The two putatively homeologous gene copies in CU samples are also highly similar except for the presence vs. absence of the 12-bp insertion, which is discussed further below.

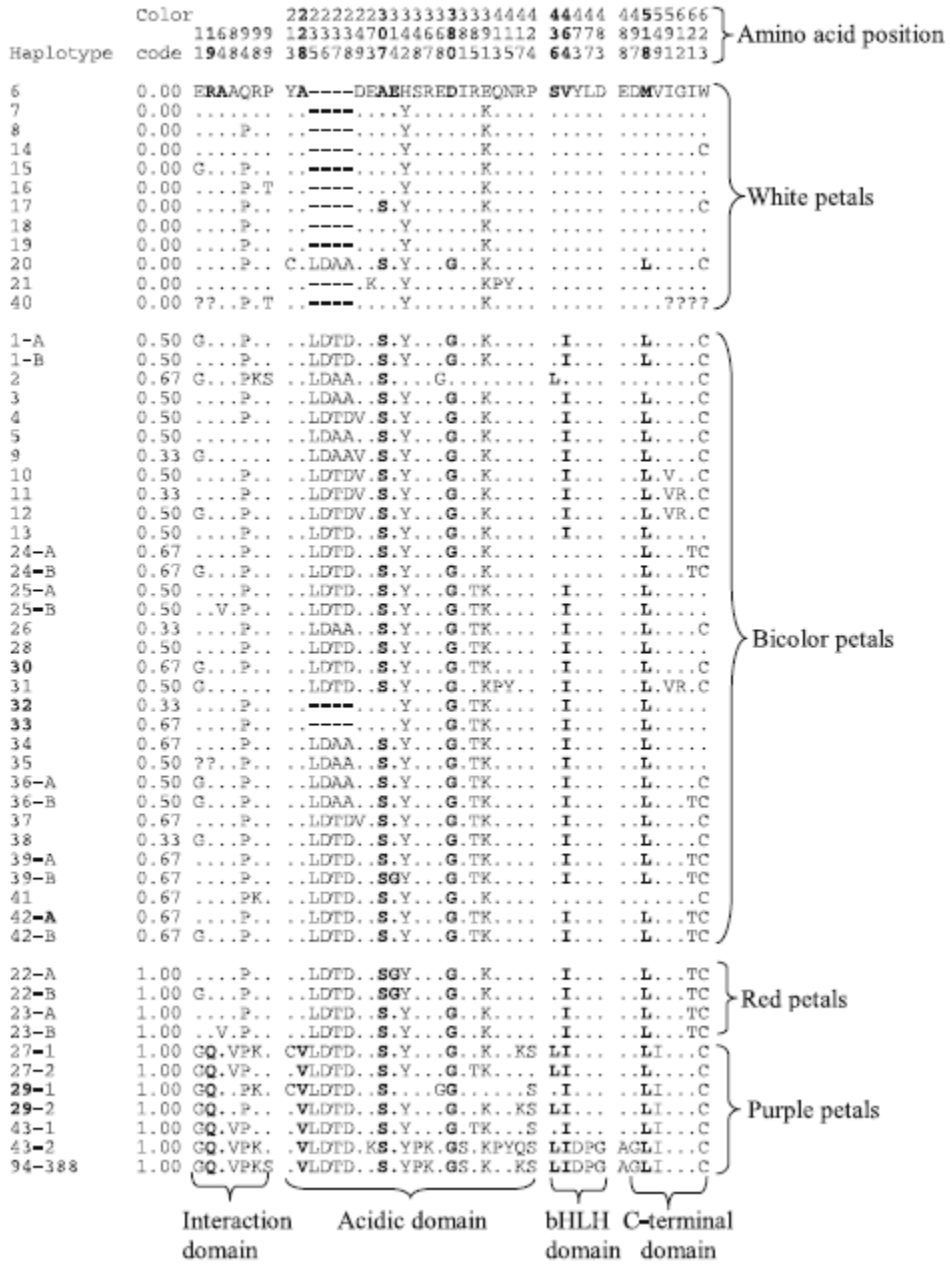

Figure 6. Data matrix of petal color scores and polymorphic amino acid sites of *antR-Cor* gene among haplotypes of the dwarf dogwoods. Dots indicate identity to topmost sequence. Dashes represent gaps. All bicolor and hybrid samples with the gap sequence presence have an allele with the gap sequences absent (not shown in the figure). ''A'' and ''B'' are designated as two alleles if a samples has two haplotypes. The seven aminoacid sites associated with petal colors are marked in bold faces.

Based on a cladistic analysis of morphological characters, Murrell (1994) divided the dwarf dogwoods into five lineages: *Cornus canadensis, C. unalaschkensis*, and *C. suecica* as distinct species, respectively, and *C. canadensis* > *C. suecica* and *C. suecica* > *C. canadensis* as two informal categories. *Cornus unalaschkensis* was considered a tetraploid derived from hybridization between two diploid species (*C. canadensis* and *C. suecica*) (Dermen 1932; Taylor and Brockman 1966; Clay and Nath 1971; Bain and Denford 1979). *Cornus canadensis* > *C. suecica* was considered the product of backcrossing of hybrids to *C. canadensis*, and morphologically more similar to *C. canadensis*. Similarly, *Cornus suecica* > *C. canadensis*, morphologically closer to *C. suecica*, were considered to be the products of backcrossing of hybrids to *C. suecica*. We found no evidence supporting five distinct lineages in the complex. Our data provide support for two distinct lineages, one consisting of *C. suecica* and the other consisting of *C. canadensis, C. unalaschkensis*, and the diploid intermediates. We found fixed molecular differences between the two diploid species that are divergent in petal colors. Moreover, our results from the *antR-Cor* sequence data appear to raise the possibility that the bicolored phenotype may not have originated from hybridization between *C. canadensis* and *C. suecica*, but may instead represent the progenitor condition of the *C. canadensis*–bicolored complex, with the white flowers of C. canadensis having arisen later. Four lines of evidence support this hypothesis. First, the root position in the neighbor-joining analysis suggests that alleles from bicolored individuals are in basal positions in the *C. canadensis*–bicolored complex, with *C. canadensis* alleles all within a derived subclade that has moderate bootstrap support of 62% (Fig. 4). Second, *C. canadensis* has the lowest level of polymorphism among the phenotypic groups identified in this study (Table 3), which may be evidence of a genetic bottleneck associated with its derivation from a bicolored progenitor population. Third, *C. unalaschkensis* and the diploid bicolored group show only slightly less sequence divergence from *C. suecica* than does *C. canadensis* (Table 4). Fourth, all alleles bicolored individuals are part of a complex and extensively recombinant network of haplotypes encompassing *C. canadensis* and *C. unalaschkensis* as well as the diploid bicolored samples but distinct from *C. suecica* (Fig. 5). Our method for inferring haplotypes in CU individuals should, if anything, have biased the results of the haplotype cluster analyses in the direction of identifying two distinct haplotype classes. Instead, the two inferred haplotypes from CU samples tended to cluster together (Fig. 4). The high degree of similarity between the two copies of *antR-Cor* in *C. unalaschkensis* is inconsistent with an allotetraploid origin, for which the expectation would be for each individual to contain one copy each from *C. canadensis* and *C. suecica*. Alternatively, it is possible that chromosomes pair randomly into bivalents in an allotetraploid *C. unalaschkensis* rather than pairing as differentiated homeologs, and that the *C. suecica* chromosomes have been lost due either to random processes or selection. However, this would not explain the similar absence of *C. suecica* alleles in all diploid CH samples with the exception of one purple-petaled individual that is likely to be a recent hybrid. It is also possible that limited sampling across the geographic range of *C. suecica* resulted in failure to find CS haplotypes within the CC/CH/CU clade. All but one of the *C. suecica* samples, however, came from areas in which CC and CH were also sampled. The interspersion of CC, CH, and CU alleles in the haplotype cluster appears more consistent with *C. unalaschkensis* being an autopolyploid derived from CH-like individuals rather than an allopolyploid hybrid of *C. canadensis* and *C. suecica* (Figs. 4, 5). Random chromosome pairing into bivalents would explain the lack of differentiation between *antR-Cor* copies in the CU samples. Under this scenario, if the frequencies of alleles with and without the

12-bp insertion were similar, tetraploids homozygous for the presence or absence of the insertion would be relatively rare, making their absence from the CU samples unsurprising. The sequence and haplotype diversity of *C. unalaschkensis*, which is only slightly less than that of the diploid intermediates, suggests either that *C. unalaschkensis* arose from a relatively large initial population of tetraploids or that tetraploids have arisen on multiple occasions.

This hypothesis of a non-hybrid origin for the bicolored complex does not preclude the occurrence of hybridization between *C. canadensis*, bicolored diploid populations, and *C. suecica*. Sample #29, which contains two *C. suecica antR-Cor* alleles (Fig. 3) and *C. suecica* flower color but has other morphological evidence of hybrid ancestry, seems likely to be the result of recent hybridization between *C. suecica* and diploid bicolored ancestors. Moreover, the range of intermediate color and morphological characteristics in the bicolored samples may result in part from hybridization between *C. canadensis* and bicolored populations.

The alternate scenario presented here for the origins of *C. unalaschkensis* and diploid intermediates must be treated only as a hypothesis at present, and evaluated in the context of morphological evidence. It is possible that the phylogenetic pattern and the lower sequence diversity observed for *C. canadensis* is due to sampling error (e.g., ancestral haplotypes of *C. canadensis* were missed in the sampling or have been lost). Moreover, the patterns found in *antR-Cor* may not be concordant with the evolution of the dwarf dogwoods complex due to incomplete lineage sorting, effects of hybridization, and possibly selection (Felsenstein 2004). Thus, hypotheses for the origins and evolution of the dwarf dogwoods complex should be tested by isolating and sequencing additional nuclear genes with sampling from other geographic regions. If our hypothesis of a functional role for *antR-Cor* is correct (see below), genes unlikely to be involved in flower color or other aspects of morphological variation should not show similar patterns of allelic diversity and relationships to those reported for the *antR-Cor* gene.

Given the widespread distribution of the ‘‘CU’’ group, it is possible that selection favors the bicolored tetraploid genotypes, which may permit adaptation to a wider range of environments. For example, bicolor petals would be beneficial for plants for attracting pollinators and improving stress tolerance, or this trait could be associated with some other advantageous phenotypes that could help tetraploids and diploid intermediates adapt to new and diverse niches.

Correlation between color of petals and gene evolution

We considered the possibility of a functional role for *antR-Cor* in flower color variation. First, there is abundant evidence that mutation in anthocyanin regulatory genes can lead to petal color changes (see review by Mol et al. 1998). Second, some polymorphisms resulting in non-conservative amino acid changes sort out with the flower color differences in the dwarf dogwood complex. While a statistical test of sequence-phenotype associations is not valid due to lack of random mating within the overall complex, it is possible that the observed sorting has a functional bias. Activities and properties of proteins are the consequence of interactions among their constitutive amino acids. Therefore, the changes of amino acids with different chemical properties (e.g. side chain, electrostatic interactions, hydrophobic effects, and the size of residue) will potentially affect the structure and function of proteins (Atchley et al. 2000).

Substitutions in four sites (sites 19, 307, 380, and 436) between two alleles involve the residues with different chemical structures and/or charges (Fig. 6). Site 19 in the interaction domain involves the substitution between arginine (white flowers), a basic amino acid, and glutamine (purple flowers). The uncharged residues in the interaction domain are believed to play an important role in forming a hydrophobic core for regulatory proteins. Glutamine as an uncharged amino acid, thus, might play a role in maintaining the normal regulatory function in generating the purple color phenotypes. Site 436 is located in the bHLH domain, and shows a substitution between serine (S) and leucine (L). Leucine is mainly found in samples with purple petals. Neither leucine nor serine is a basic amino acid, but they differ in their hydrophobicity. Leucine, a hydrophobic residue, is required for dimer formation. Serine, in contrast, has a hydrophilic side chain that may be required for binding in the basic region. Thus, substitutions between leucine and serine may cause improper binding. In the acidic domain, our data show that substitutions in two sites involving amino acids with different chemical properties (site 307-alanine vs. serine, and site 380-glysine vs. aspartic acid) are associated with petal color. At site 307, alanine, a hydrophobic residue, is found only in samples with white petals, with two exceptions, and the presence of serine, which is hydrophilic, is associated with purple petals. At site 380, aspartic acid, an acidic and hydrophilic amino acid, is nearly fixed with white petal phenotype, and the presence of glycine, a non-acidic and hydrophobic amino acid is associated with purple petals. This substitution at site 380 seems to counter the expectation that in the acidic domain; acidic/hydrophilic amino acids are required for transactivation. However, glycine is a very small amino acid; its replacement by the large aspartic acid may disrupt the protein secondary structure, potentially effecting the transactivation in the white flower.

The observed relationship between gene sequences and petal color phenotypes in the hybrids indicate that the *antR-Cor* gene could be at least partially responsible for the petal color patterns in the dwarf dogwoods. Over the long term, mutations in the anthocyanin regulatory gene may have led to species divergence in the dwarf dogwoods. None of the sites we have identified show a complete association with flower color, but it is possible that the *antR-Cor* gene controls flower color in combination with genes at other loci. Alternatively, sites in regulatory regions outside the coding region may be responsible for functional differences in the *antR-Cor* alleles.

Our results raise new questions about the nature of intermediate phenotypes within the dwarf dogwood complex, suggesting that they may represent an ancestral condition rather than the results of more recent hybridization. Nevertheless, there is evidence for ongoing hybridization, which may be responsible for some of the phenotypic variability within the complex. It is well documented that hybridization followed by introgression may lead to transfer of traits from one taxon into another, allowing for range expansion of the introgressed form (e.g. Lewontin and Birch 1966). Similarly, polyploidy events such as those that gave rise to *C. unalaschkensis* can create novel opportunities for adaptive evolution due to the extensive opportunities for functional diversification in duplicated genes (Wendel 2000). Both hybridization and polyploidy are evident in the evolution of the dwarf dogwoods, and may have contributed to variation in pigmentation and consequent range expansion in the complex. Resolving the evolutionary relationships within the complex and determining the molecular basis for phenotypic evolution will require more detailed functional analysis of the *antR-Cor* gene, combined with analysis of sequence variation at additional loci.

Acknowledgments

The authors thank the following people for their help with the study: Brian Cassel for assistance with sequencing; Jingen (Jim) Qi, Christian Brochmann, Margaret Ptacek, and Jean Schulenberg for plant sample collection; Nina Gardner for DNA extraction and morphological identification; members of the Xiang lab for a variety of help and discussion; Becky Boston for using her lab space in the experiments; and, Tom Wentworth and two anonymous reviewers for critically reading the manuscript. This study is supported by Faculty Research Grants from Idaho State University and North Carolina State University and NSF grant DEB-0129069 to Q.-Y.X., and Karling Graduate Student Research Award from Botanical Society of America and Deep Gene Travel Award from Deep Gene Research Coordination Network (NSF DEB-0090227 funded to B. D. Mishler) to C.F.

References

Atchley WR, Wollenberg KR, Fitch WM, Terhalle W, Dress AW (2000) Correlation among amino acid sites in bHLH protein domains: an information theoretic analysis. Mol Biol Evol 17:164–178

Bain JF, Denford KE (1979) The herbaceous members of genus *Cornus* in NW North America. Bot Notiser 132:121–129

Carroll SB (1995) Homeotic genes and the evolution of arthropods and chordates. Naturalist 376:479–485

Clay SN, Hath J (1971) Cytogenetics of some species of *Cornus*. Cytologia 36:716–730

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9:1657–1659

Consonni G, Viotti A, Dellaporta SL, Tonelli C (1992) cDNA nucleotide sequence of *Sn*, a regulatory gene in maize. Nucleic Acids Res 20:373

Consonni G, Geuna F, Gavazzi G, Tonelli C (1993) Molecular homology among members of the *R* gene family from maize. Plant J 3:335–346

Dermen H (1932) Cytological studies of *Cornus*. J Arnold Arboretum 13:410–417

De-Vetten N, Quattrocchio F, Mol J, Koes R (1997) The *an1* locus controlling flower pigmentation in *Petunia* encodes a novel WD-repeat protein conserved in yeast, plants, and animals. Genes Dev 11:1422–1434

Dickinson WJ (1988) On the architecture of regulatory systems: evolutionary insights and implications. BioEssays 8:204–208

Doebley J (1993) Genetics, development and plant evolution. Curr Opin Genet Dev 3:865–872

Doebley J, Lukens L (1998) Transcriptional regulators and the evolution of plant form. Plant Cell 10:1075–1082

Fan C, Xiang Q-Y (2001) Phylogenetic relationships within *Cornus* L. (Cornaceae) based on 26S rDNA sequences. Am J Bot 88:1131–1138

Fan C, Purugganan MD, Thomas DT, Wiegmann BM, Xiang Q-Y (2004) Heterogeneous evolution of the *myc*-like anthocyanin regulatory gene and its phylogenetic utility in *Cornus* L. (Cornaceae). Mol Phylogen Evol 33:580–594

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Felsenstein J (2004) Inferring phylogenies. Sinauer, Sunderland, MA

Ferris S, Whitt G (1979) Evolution of the differential regulation of duplicate genes after polyploidization. J Mol Evol 12:367–317

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Goldsbourougn AP, Tong Y, Yoder JI (1996) *Lc* as a non-destructive visual reporter and transposition marker gene for tomato. Plant J 9:927–933

Gong Z, Yamagishi E, Yamazaki M, Saito K (1999) A constitutively expressed *myc*-like gene involved anthocyanin biosynthesis from *Perilla frutescens*: molecular characterization, heterologous expression in transgenic plants and transactivation in yeast cells. Plant Mol Biol 41:33–44

Goodrich J, Carpenter R, Coen ES (1992) A common gene regulates pigmentation pattern in diverse plant species. Cell 68:955–964

Gunatilleke CVS, Gunatilleke AUN (1984) Some observations on the reproductive biology of three species of Cornus (Cornaceae). J Arn Arb 65:419–427

Hu J, Anderson B, Wessler SR (1996) Isolation and characterization of rice *R* genes: evidence for distinct evolutionary paths in rice and maize. Genetics 142:1021–1031

Kellogg EA (2002) Root hairs, trichomes and the evolution of duplicate genes. Trends Plant Sci 6:550–552

King JL, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. Science 188:107–116

Levine M, Tjian R (2003) Transcription regulation and animal diversity. Nature 424:147–151

Lewonton RC, Birch LC (1966) Hybridization as a source of variation for adaptation to new environments. Evolution 20:315–336

Lioyd AM, Walbot V, Davis RW (1992) *Arabidopsis* and *Nicotiana* anthocyanin production activated by maize regulators *R* and *C1*. Science 258:1773–1775

Ludwig S, Wessler SR (1990) Maize *R* gene family: tissue-specific helix-loop-helix proteins. Cell 62:849–852

Martin C, Prescott A, Mackay S, Barlett J, Vrijlandt E (1991) Control of anthocyanin biosynthesis in flower of *Antirrhinum majus*. Plant J 1:37–49

Mol J, Grotewold E, Koes R (1998) How genes paint flowers and seeds. Trends Plant Sci 3:212–217

Moore RC, Purugganan MD (2005) The evolutionary dynamics of plant duplicate genes. Curr Opin Plant Biol 8:122–128

Murrell ZE (1994) Dwarf dogwoods: intermediacy and the morphological landscape. Syst Bot 19:539–556

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, USA

Palopoli MF, Patel N (1996) Neo-Darwinian developmental evolution- can we bridge the gap between pattern and process?. Curr Opin Genet Dev 6:502–508

Papp B, Pál C, Hurst LD (2003) Evolution of *cis*-regulatory elements in duplicated genes of yeast. Trends Genet 19:417–422

Purugganan MD (1998) The molecular evolution of development. BioEssays 20:700–711

Purugganan MD (2000) The molecular population genetics of regulatory genes. Mol Ecol 9:1451–1461

Purugganan MD, Suddith JI (1998) Molecular population genetics of the *Arabidopsis CAULIFLOWER* regulatory gene: nonneutral evolution and naturally occurring variation in floral homeotic function. Proc Natl Acad Sci USA 95:8130–8134

Purugganan MD, Suddith JI (1999) Molecular population genetics of floral homeotic loci: departures from the equilibrium-neutral model at the *APETALA3* and *PISTILLATA* genes of *Arabidopsis thaliana*. Genetics 151:839–848

Quattrocchio F, Wing JF, Leppen HTC, Mol JNM, Koes RE (1993) Regulatory genes controlling anthocyanin pigmentation are functionally conversed among plant species and have distinct sets of target genes. Plant Cell 5:1497–1512

Quattrocchio F, Wing JF, Woude KVD, Mol JNM, Koes RE (1998) Analysis of bHLH and MYB domain proteins: species-specific regulatory differences are caused by divergent evolution of target anthocyanin genes. Plant J 13:475–488

Radicella PD, Turks D, Chandler VL (1991) Cloning and nucleotide sequence of a cDNA encoding *B-Peru*, a regulatory protein of the anthocyanin pathway in maize. Plant Mol Biol 17:127–130

Riechmann JL, Meyerowitz EM (1997) MADS domain proteins in plant development. Biol Chem 378:1079–1101

Rozas J, Sánchez-Delbarrio JC, Messeguer X, Rozas R (2003) DNASP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Simpson P (2002) Evolution of development in closely related species of flies and worms. Nature Rev Genet 3:907–917

Stapleton A (1992) ultraviolet radiation and plants: burning questions. Plant Cell 4:1353–1358

Swofford DL (2002) PAUP: phylogenetic analysis using parsimony, version 4.0b10. Sinauer Associates, Sunderland, MA

Tajima F (1989) Statistical methods for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:595–595

Taylor RL, Brockman RP (1966) Chromosome numbers of some western Canadian plants. Can J Bot 44:1093–1103

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III Cladogram estimation. Genetics 132:619–633

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Watterson G (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276

Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol 42:225–249

Wilson AC (1975) Evolutionary importance of gene regulation. Stadler Symposium vol 7. University of Missouri, Columbia, Missouri, pp 117–134

Wolfe KH (2001) Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet 2:333–341

Xiang QY, Soltis DE, Soltis PS (1998) Phylogenetic relationships of Cornaceae and close relatives inferred from *matK* and *rbcL* sequences. Am J Bot 85:285–297

Xiang Q-Y(J), Thomas DT, Zhang WH, Manchester SR, Murrell Z (2006) Species level phylogeny of the Dogwood genus *Cornus* (Cornaceae) based on molecular and morphological evidence – implication in taxonomy and Tertiary intercontinental migration. Taxon 55:9–30